# ) TENT COOPERATION TREETY

<del>r</del>	I To:

## **PCT**

### **NOTIFICATION OF ELECTION**

(PCT Rule 61.2)

1	U	11	ı	LI I	E	11	V		ור	V	~	١	! !	C	"	V	4	L	D	U	П	/

Assistant Commissioner for Patents United States Patent and Trademark Office

Box PCT

Washington, D.C.20231 ÉTATS-UNIS D'AMÉRIQUE

Date of mailing (day/month/year) 23 November 1999 (23.11.99)	in its capacity as elected Office
International application No. PCT/CN99/00046	Applicant's or agent's file reference P199103E
International filing date (day/month/year) 06 April 1999 (06.04.99)	Priority date (day/month/year) 06 April 1998 (06.04.98)
Applicant	
LIU, Sha	

		··		
1.	The designated Office is hereby notified of its election made:			
	X in the demand filed with the International Preliminary Examining Authority on:			
l	29 October 1999 (29.10.99)	_		
	in a notice effecting later election filed with the International Bureau on:			
		-	,	
2.	The election X was			
-	was not	•	•	-
	made before the expiration of 19 months from the priority date or, where Rule 32 app Rule 32.2(b).	lies, within the tin	ne limit under	

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

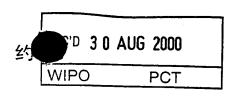
Authorized officer

Ting Zhao

Facsimile No.: (41-22) 740.14.35

Telephone No.: (41-22) 338.83.38

# 专利合作条 PCT



# 国际初步审查报告

(PCT 条约 36 和细则 70)



中请人或代理人的档案号 PI99103E	关于后续行为	参见"传送国际	示初步审查报告的通知"(PCT IPLA 416 表)
国际中请号	国际中请日(日/月	/年)	优先权日(日/月/年)
PCT/CN99/00046	06.4 月 199	9(06.04.99)	06.4 FJ 1998(06.04.98)
国际专利分类(IPC)或者国家分类和 G06F 17/28	IPC 两种分类		
申请人 刘 莎			
<ol> <li>本国际初步审查单位已作出国</li> <li>本报告共计3页,包括扉页。</li> <li>本报告还有附件,即修改页,和/或对本国际初步审这些附件共计页</li> </ol>	后的并且作为本报告表	表础的说明 书修	6改页、权利要求书修改页和 或副图等:
3. 本报告包括关于下列各项的内容 1 図 报告的基础	ř:		
II □ 优先权			
III	创造性和工业实用性的	意见	
IV D 缺乏发明的单一性			
v 🗵 按条约 35(2)关于新颖	性、创造性或工业实质	用性的推断性意	意见:支持这种意见的引证和解释
VI U 引用的某些文件			
VII U 国际中请中的某些缺陷	14 60		
VIII			
提交要求书的日期		完成本报告的	日期
29.10 月 1999(29.10	0.99)		15.8月 2000(15.08.00)
国际初步审查单位名称和地址 IPEA/CN 中国北京市海淀区西土城路 传真号:86-10-62019451	6号(100088)	受权官员 夏 电话号码: 86-	红夏红红夏印国



	你中请号
T	PCT CN99 00046

1. 报告的基础				
1. 关于国际中请中各	个部分: *			
☑ 原始提交的□	国际中请。			
说明书,	郊	页.按 原始提交的,		
	第	页, 随 要求书提交的	勺.	
_		页, 随		_ 的信件提交的
□ 权利要求,	郊	页,原 始提交的,		:
	第	页,按 条约第 19 条	修理工改的(附有说明)。	
	郊	页,随 要求书提交的	勺。	;
		页, 随		] 的信件提交的 [1]
□ 附图.	第页,			:
	第页,	随要求书提交的,		
		<u> </u>	—————的信件提交的	的 ·
□ 说明书中的	序列表部分			i
	郊	页,原始要求提交 页,随要求书提交	钓,	
	郊	页,随要求书提交	钓,	
	郊	页, 随		的信件提交的
				!
这种语言是 □ 为了国  □ 本国际	际检索而提交的译本 中请公布时所使用的	國际初步审查单位提交的这 所使用的语言(细则 23.1 语言(细则 48.3 (b))。 译本所使用的语言(细则 5	(b)) .	
□ 国际中记 □ 与国际以 □ 与国际以 □ 后来以 □ 已提交 □ 已提下 □ 已以 I 下 □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □	青中所包含的书写形中请同时提交的计算 书写形式向本国际初计算机可读的形式向 了关于后来提交的书 了关于以计算机可读 1容的:	机可读形式的序列表。 步审查单位提交的序列表。 本国际初步审查单位提交的 写形式的序列表没有超出房 的形式记载的信息是与书写	]序列表。 - 始提交的国际中请所公+	F的范围的说明:
□ 权利要求 □ 附图,	, 筑页 筑页	页 		
5. □ 由于(某些) 出的(细则 70.		始公开的范围,如补充栏所。	5,因此本报告是按照如同	没有修改的情况作
本报告的附件, 因为	它们没有包含修改(组	的替换页,在本报告中被称为 II则 70.16 和 70.17 )。 1 项中指明,并作为本报告的	٠.	· <b>贝</b> 不作为

性条件.



### 国际初步审查报告

国际申请号

PCT CN99 00046

意见		
新颖性(N)	权利要求 1-14	Į.
	权利要求	11
创造性(IS)	权利要求 1-14	<u>}</u> _
	权利要求	11
工业实用性(IA)	权利要求 1-14	是
	权利要求	11



### INTERNATIONAL PRELIMINARY EXAMINATION REPORT

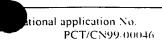
(PCT Article 36 and Rule 70)

Applicant's or agent's file reference PI99103E	FOR FURTHER A		of Transmittal of International Preliminary ort (Form PCT/IPEA-416)
International application No.	International filing d	ate (day/month/year)	Descript data (d
PCT/CN 99/00046		1999(06.04.99)	Priority date (day/month year)
International Patent Classification (IPC) or			06 April.1998(06.04.98)
G06F 17/28	national classification	and IFC	
6001 17/20			
Applicant			
LIU,Sha			
This international preliminary exami	ination report has beer	prepared by this Internation	onal Preliminary Examining Authority and
is transmitted to the applicant according	•		,
2. This REPORT consists of a total of	-	sheets, including this	s cover sheet.
This report is also accompanied by	ANNEXES, i.e., she	eets of the description, cla	aims and /or drawings which have been
amended and are the basis for this rep	ort and/or sheets conta	ining rectifications made be	efore this Authority ( see Rule 70.16 and
Section 607 of the Administrative Inst	ructions under the PCT	).	
These annexes consist of a total of		sheets.	
-			
3. This report contains indications rela	iting to the following it	ems:	
1 Basis of the report			
priority			
III Non-establishment of opinion	on with regard to novel	ty,inventive step and indust	trial applicability
I\' Lack of unity of invention			
V Reasoned statement under A			or industrial applicability:
citations and explanations su	apporting such statemen	11	
VI — Certain documents cited			
Certain defects in the interna	ational application		
Certain observations on the	international opplication	on.	
Date of submission of the demand		Date of completion of this	s report
29 October. 1999(29.10.	.99)	15 A	ugust.2000(15.08.00)
Name and mailing address of the IPEA/CN	V	Authorized officer	4- 百
6 Xitucheng Rd., Jimen Bridge, Haidian D			Guohong, Xia
100088 Beijing, Chin Facsimile No. 86-10-62019451	a	Telephone No.86-10-620	93837
B B B B B B B B B B B B B B B B B B B		·	

Form PCT/IPEA/409(cover sheet)(July 1998)

l. E	Basis of t	he report		
	regard	to the elements of the international application:*		
	the int	ernational application as originally filed		
	the de	cription:		
	pages			as originally filed
	pages			filed with the demand
	pages		,filed with the letter of	
	the clo	ims:		
	pages			as originally file.
	pages		, as amended (together v	with any statement)under Article 19
	pages			.filed with the demand
	pages		,filed with the letter of	
	the des	cription:		
	pages			as originally filed
	pages			filed with the demand
	pages		,filed with the letter of	_
	the sec	uence listing part of the description:	,	
	pages			as originally filed
	pages		- M	filed with the demand
	pages		,filed with the letter of	
The	the lar the lar the lar the ba and/or		e following language	examination (under Ruls Rules 55/2)
3. With prelii		to any nucleotide and/or amino acid sequence camination was carried out on the basis of the sequence		ional application, the international
		d in the international application in written form.		
		ether with the international application in computer red subsequently to this Authority in written form.	adable form.	
		I subsequently to this Authority in computer readable	form.	
	The stat	ement that the subsequently furnished written sequence tion as filed has been furnished.		nd the disclosure in the international
	The sta	tement that the information recorded in computer readed.	dable form is identical to t	the written sequence listing has been
4.	The amo	the description,pages the claims Noa. the drawings,sheets/fig		
5. Th	nis repor	has been established as if (some of )the amendmen	ts had not been made, sin	nce they have been considered to go
_		isclosure as filed, as indicated in the Supplemental Be		
•	his repoi	heets which have been furnished to the receiving Office t as "originally filed" and are not annexed to this r	=	
** Any re	placeme	nt sheet containing such amendments must be referred	to under item I and anne:	xed to this report.





citations and explanations supp		gard to novelty, inventive step or industrial applical tatement	•
Statement:			
Novelty (N)	Claims	1-14	YES
Novelly (IV)			
	Claims		NO
Inventive step (IS)	Claire	1.14	X11.
inventive step (15)	Claims	· · · · · · · · · · · · · · · · · · ·	
	Claims		
Industrial applicability (IA)	Claria.	1.14	<b>X</b> *1 · · ·
musicial application (171)	Claims		
	Claims		\()

# 专利合作条约

A READE 2000

发信人: 国际初步审查单位

收借人:

#### 100032

北京金融大街 27 号投资广场 A 座 10 楼 永新专利商标代理有限公司 北京办事处

## PCT

传送国际初步审查报告通知书

(PCT 细则 71.1)

17

发文日(日/月/左4. 8月 2000 (24.08.00)

 申请人或代理人的档案号
 重要通知

 PI99103E
 重要通知

 国际申请号
 国际申请日(日/月/年)
 优先权日(日7月/年)

 PCT/CN99/00046
 06.4 月 1999(06.04.99)
 06.4 月 1998 (06.04.98)

刘沙

- 1. 通知申请人,本国际初步审查单位随本通知传送对国际申请制定的国际初步审查报告及其附件(每果了) 附件的话)。
- 2. 报告及其附件(如果有附件的话)的副本同时送交国际局,以便送达所有选定局。
- 3. 任何选定局提出要求时、国际局将作出报告的英文译文(但不是任何附件的译文),并将该译文()。 给这些选定局。
- 4. 提示

在自优先权目起 30 个月内(或者在有些局更迟)申请人必须完成一定的行为(提交还本和意义 家费)进入各选定局的国家阶段(条约第 39 条 (1))(参见国际局寄送的 P C T | 1 B 301 百年 + 的提示)

国际申请的译本必须向选定局提供时,该译本还必须包括国际初步审查报告两件的译文。等按向各有关选定局提供该译文是申请人的责任。

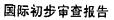
有关各选定局适用的期限和要求的详情,参见PCT申请人指南第11卷

71.				
国际初步审查单位名称和地址 IPEA/CN 100088 中国北京市海淀区蓟门桥西上城路 6 号	受权官员。		्रा <sub>क</sub>	 
11 11 11 11 11 11 11 11 11 11 11 11 11	कार्य है ।	62093837		

# 专利合作条约 PCT

## 国际初步审查报告 (PCT 条约 36 和细则 70)

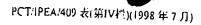
中请人或代理人的档案号 PI99103E	关于后续行为 参见"传送国	际初步审查报告的通知"(PCT IPFA 416
国际中请号 PCT/CN99/00046	国际中请日 <i>(日/月/年)</i> 06.4 月 1999(06.04.99)	优先权日(日/月/年) 06.4 月 1998(06.04.98)
国际专利分类(IPC)或者国家分类和 IPC G06F 17/28	C 两种分类	
中请人 刘 莎		
		<b>8改页、权利要求书修改页和</b> 或第译 [4]
3. 本报告包括美于下列各项的内容:  □		《见:支持这种意见的引证和解释
是交要求书的日期 29.10 月 1999(29.10.99) 母院 初步軍責单位名称和地址 IPEA/CN 中国北京市海淀区西土城路 6 号	受权官员 夏	15.8 J) 2000(15.08.00)



国际	41	语号

PCT/CN99/00046

			i	1 C 1/ C 1499/00040
1. 报告的基础	_			
1、关于国际中请中名	≤个部分:*			,,,,,
│ │ │ │ │ │ │ │ │ │ │ │ │ │ │ │ │ │ │	国际中请。			
رى رواين 🔲 ا		页,按原	产料态的	
	ar			
				14 10 11 11
□ 权利要求。	27		A AA	
12.17.2.30	郊 郊			1111111
	រា:		的第 19 条修理 11改的(附	有说明)。
│ □ 胸圏、	郑页, ]	火,飓 点必根		
	郑页,			
				And the transfer
०० का मा धारण	字列表部分		<del></del>	信件提交的.
ייין מייטע <u>שר</u> 		ਦ <i>ਅ</i> :///	m -2-12	
	第 第	页,原始		
		页. 随要:		
	45	贝. 随		
□ 本国际□	际检索而提交的译本原 中请公布时所使用的语 家初步审查而提公的语	善言(邹则 48.3(		
3. 关于本国际申请中	所公开的任何核甙酸	和/或氨基酸的序		。 模据下面的序列表过有的:
	青中所包含的书写形式			
	中请问时提交的计算机			
	5′写形式向本国际初2			
	上算机可读的形式向本			
□ □ □ □提交〕	『关于后来提交的书》	形式的序列表没	有超出原始提交的国际中	1请所公开的范围的说识。
□ □ □提交〕	" 关于以计算机可读的	的形式记载的信息。	是与书写形式的序列表框	1司的说明。
4. 錄改删除了以下內	容的:			
□ 说明 15.	<b>氮</b>	页		
── 权利要求.	郊			
上」脚图.	第页	. 函		
5. 日 出了(某些)。 出的(細頭 70.2		公开的范围,如补	充栏所示。因此本报告是	按照如同没有你饮作。。
			中破称为"原始提交的"。	这些特殊的不少
	它们没有包含橡改(细)			
** 任何包含这种橡皮的	替换页,都必须在第十.	威中指明, 并作为才	(报告的阳件)	



# 国际初步审查报告

1	国际中错号
1	PCT/CN99/00046

	拉名约 35 冬(3) 芒	T OF ETALL Address of the second	
_	18年27 33 宋(2)天	于新颖性、创造性或工业实用性的推断性意见、支	持这种意见的引证和解释
	意见		
	新颖性(N)	权利要求 <sub>1-14</sub>	是
		权利要求	
	创造性(IS)	权利要求 1-14	是
		权利要求	
	T.业实用性(IA)	枚利要求 1-14	是
		权利要求	

2. 引征和解释(组则70.7)

在检索中未发现影响该中请新颗性。创造性的对比文件,当前权利要求 1-14 具备获得专利权的新颗性和创造性条件。

^	4	
(Deg.	4	
1 1-0	7.0	IVI
11/201	u es c	
1 - 1		
' \		

# **PCT**

# 请 求 书

下列签字人请求按照

————由 易	_
国际申请号	
国际中请日	
受理局名称和 'PCT 国际申请'	

专利合作条约的规定处理本国际申请	申请人或代理人的档 (如果有)(压12个字符	· · DIODIOX#
第1栏 发明名称 开放式全息模板式人	机对话语言翻译	<b>P</b> 方法及全息语义标注系统
第Ⅱ 栏 申请人		
姓名 (或名称)和地址: (姓在前,名在后;法人应该写正式会 码和图名。如果下面来指明居所,则本栏中指明地址的所属		囚 该人也是发明人
# 4 <b>本</b> 4 )		电话号码:
LIU, Sha 中国北京市海淀区牡丹园东里 10 号楼]	1307 室 <i>(</i> 100083)	传真号码:
Room 1307,Building No.10, Mudanyud Haidian District, Beijing 100083,P.R.	an Dong Li, China	电传号码:
<b>國籍 (國本名称):</b>	居所(国家名称):	CN
该人是对下列国家的申请人:图所有指定国 □除美	国以外的所有指定国	□ 美国 □ 补充栏中注明的国家
第Ⅲ 栏 其他申请人和/或(其他)发明人		
姓名 (或名称)和地址: (姓在前,名在后; 法人应填写正式:	全称,地址应包括邮政抵	该人是
码和国名。 如果下面未指明居所, 则本栏中指明地址的所属 家名称))	国为申请人的房所(国	□中请人
		□申请人和发明人
		□ 发明人(如果论择此方格,不必填写以下诸項。)
国籍(国家名称):	居所(回家名称):	
该人是对下列国家的申请人:□所有指定国 □ 除》	<b>美国以外的所有指定国</b>	□ 美国 □ 补无柱中注明的国家
□ 共余中请人和/ 或发明人注明在续页中		
第 IV 栏 代理人或共同代表;或通信地址		
下列人员被委托/ 已经被委托作为申请人向主营国际单位办	7理事务的: 囚 代理人 [	□共同代表
姓名 (或名称)和地址:(姓在前,名左后;法人应该写正式会 码和国名。)	全称, 迪兰庄包括邮政站	电话号码:86-10-6621 1834-1838
永新专利商标代理有限公司 北京办事处 NTD PATENT & TRADEMARK AGENCY LTD.	BELJING OFFICE	传真号码:86-10-6621 1845-1848
北京金融大街 27 号投资广场 A 座 10 楼	(100032)	
10 <sup>th</sup> Floor, Block A, Investment Plaza, 27 Jinro Beijing 100032 P.R.China	ongdajie,	电传号码:
□通信地址: 如果永委托/未委托过代理人或共同代表,并	把上栏中运明的地址作为	)通信的专门地址,在此方格作出标记。

### 第Ⅴ栏 国家的指定

按照细则 4.9(a)规定指定下列国家或地区(在相应的方格中作出标记,至少选择一项):

### 地区专利

- 区 AP ARIPO 专利: GH 加纳, GM 赞比亚, KE 肯尼亚, LS 莱索托, MW 马拉维, SD 苏丹, SZ 斯威士兰, UG乌干达, ZW 津巴布韦, 以及任何同时是哈拉雷协定缔约因和 PCT 缔约国的国家、
- 区 EA 欧亚专利: AM亚英尼亚·AZ阿塞邦疆, BY白俄罗斯·KG古尔吉斯斯坦·KZ哈萨克斯坦, MD摩尔多瓦共和园, RU俄罗斯联邦·TJ塔吉克斯坦, TM 土库曼斯坦, 以及任何同时是欧亚专利公约缔约国和PCT缔约国的国家。
- 区 EP 欧洲专利: AT 奥地利, BE 比利时, CH 和 LI 瑞士和列支敦士程, CY 塞浦路斯, DE 德园。DK 丹麦, ES 西班牙, FI 芬兰, FR 法园。GB 英园。GR 希腊, IE 爱尔兰, IT 意大利, L U卢森堡、MC 摩纳哥, NL荷兰, PT葡萄牙, SE 瑞典, 及任何同时是欧洲专利公约缔约国和 PCT 缔约园的国家。
- 図 OA OAPI专利: BF布基纳法索,BI贝宁、CF中非共和国,CG刚果,CI科特迪瓦、CM 喀麦隆、GA加蓬、GN几内亚、GW 几西亚比绍,ML 马里,MR 毛里塔尼亚、NE 尼日尔、SN 塞内加尔,TD乍得,TG 多哥,及任何同时是非洲知识产权组织缔约国和 PCT 缔约国的国家(如果要求得到其他保护或待遇、请在虚线上注明)、

			. nn 1		
家	专利	(如果要求得到其他保护或待遇,请在虚线上注	E 193 J	1	A
X	ΑL	阿尔巴尼亚	X	LR	利比里亚
IX)	AM	亚美尼亚	X	LS	莱宏托
K		奥地利	X	LT	立陶宛
K	ΑŬ	澳大利亚	X	LU	卢森堡
X		阿塞拜疆	<b>X</b>	LV	拉脱维亚
Z)		波斯尼亚和黑塞哥维那	X	MD	<b>摩尔多瓦共和国</b>
X		巴巴多斯	X	MG	马达加斯加 · · · · · · · · · · · · · · · · · ·
<b>2</b> 20	RG.	保加利亚	X	MK	前南斯拉夫马其顿共和国
[X]	RR	巴西			
_		白俄罗斯	図	MN	「 <b>蒙</b> 古
X		. 加 拿大	X	MW	/ 马拉维
X			X	мх	、墨西 丏
X	CH	和 LI 瑞士和列支敦士登	(XI	NO	挑威
X	CN	中国	121	NZ	新西兰
X	CU	古巴	[ <u>Z</u> ]	Dτ	波兰
X	CZ	捷克共和国		PT	葡萄牙
X	DE	徳国	_		罗马尼亚
X	DK	丹麦	-	KU	「俄罗斯联邦
$\mathbf{X}$		爱钞尼亚	X		
Z.		西班牙	8		苏丹
(2)	FI		X		
X		3 英因	X	SG	新加玻
X	GL	) 格林纳达 3. 格·鲁吉亚	X	SI	斯洛文尼亚
(Z)		I 加纳	X	SK	斯洛伐克
	GN	( 粉化亚	X	SL	塞拉利昂
(X)	CI	V 几内亚比绍	X	TJ	塔吉克斯坦
_		2. 克罗地亚	X		AI 土库曼斯坦
S.	111	」 匈牙利		TR	2. 土耳其
			X	] T]	「特立尼达和多巴符
[X]		中度尼西亚 以色列	Z	] UA	A 乌克兰
X			X	1 110	> 乌干汰
		冰岛 日本	[3	US	5 美国
X	JP	5 肯尼亚	ß	מוז ל	2 女炎别点斯坦
K	K	5 首定业	ß	1 VI	V 校亩
N X	K	3 吉尔吉斯斯坦	Ľ	1 71	D 南斯拉夫 ······
X	KI	9 朝鲜民主主义人民共和国	ĮX	7 71	· ************************************
ـ ا			-		
I	J KE	( 韩国	出	下谷	方格用于指定本农公布之后加人 PCT的国家(
K		7. 哈萨克斯坦		家专	111
L CX		3. 全声西亚 5. 化显光化			1 南非
<u>                                    </u>	ı Li	( 斯里兰卡 近京末羽·张以上裕定外、中诗人还法照与剧 4.9(b) 裕之张以	ør PC	T所力	, 许治定的所有 日军.
1 75	12	(a.vn. au · No M. Ca.f. & Jr & M( / L. 20) (5) (0) 24 71 72 (1) 74 74 77 (1)			E T 3 C 2 C 3 C 3 C 1 C 1 C 2 C 2 C 2 C 2 C 2 C 2 C 2 C 2

预防性指定声明: 除以上指定外, 中诗人还绘照如剧 4.9(b) 指定除以外 PCT 所允许指定的所有回答。 中诗人声明:这些补充指定是以要认为条件,并且在自优完权日起 15 个月期现局满葡萄衣 被求认的任何指定权认为被申请 人在该兵限局调讨接回。(为项认指定,有提交一份包括有美指定的通告并数的指定督和兼认货,列认应在三进 15 个月内这位受理局。)

	第 3 页		
第 VI 栏 优先权要求		其他优先权要求注目	明在补充栏中
E先申请的申请日 在先申请的申请号		在先申请是:	
(日/月/年)	国家申请:	地区中访:** 地区专利局	国际申请: 受理局
1) 6.4 月 . 1998 98101156.X	CN		
2)(20.11 月 .1998 .98125015.7	CN		
3清受理局准备并向国际局送交与上面第 情的受理局)。 如果在先申请是一项 ARIPO 申请、必须在利 (知则4.10(b)(ii))。见补充性。	1,2_项的在免申请一致		
第11 栏 国际检索单位			The second
国际检索单位(ISA)的选择 (如泉两个成者多个国际检索单位是主管 进行国际检索的单位, 请填写所选择的单位,可使用两个字母的代码来表示): ISA/ CN	病水使用任元世 (如果在先检索己语 位请求): 日期(日/月/年)	索的结果;在先检索的 由国际检索单位进行或 号码	已向国际检索学 国家(或地区专利局)
清 求 书: 3 页 说 明 书: 16 页 (除序列表部分)	·国际申请还包括下列标 1.图 费用计算页 2.图 委托书 3.□ 总委托书副本;如存 4.□ 总委签字的解释 项码 6.□ 国系子被由的公录文(它 7.□ 关于板内的或形式的 8.□ 计算它(明确指出)	「的话,登记号 注明的优先权文 音言): 生物材料保藏的单独说	明
- J1	日和捕要一起公布. 扫	是交国际申请的语言	: 中文
第1X 栏 申请人或代理人签字或意在每一签字旁注明签字人姓名,如果从请求新专利商标代理者	· · · · · · · · · · · · · · · · · · ·	还要注明此人是以什么	《名义签字的。
1. 据称的国际申请文件的实际收到日期 3. 由于随后(但在获限内)收到补充国际	中		2. 附图 □ 妆到:
访的文件或附图, 更改的实际收到日 4. 收到(在期限内)根据 PCT等	<del>胡</del> :		□ 未收至

由 受	理局填写	'
1.据称的国际中请文件的实际收到日期:		2. 附图
3.由于随后(但在疑限内)收到补充国际申 访的文件或附图,更改的实际收到日期:		<ul><li>□ 收到:</li><li>□ 未收到:</li></ul>
4.收到(在期限内)根据 PCT第 11(2)条进行的改正的日期:		
5. 国际检索单位: ISA/ (如果有两个或多个主管单位)	6. □检索本的送交拾迟到缴纳检查	<b>奈贺后</b>

国际局收到登记本的日期:

PCT/RO/101 表(末页)(1998年7月; 1999年1月再版)

**参见请求书表格的说明** 

### PCT.

# 世界知识产权组织



# 按照专利合作条约(PCT)所公布的国际申请

(51) 国际专利分类号: WO99/52041 A1 (43) 国际公布号: WO99/52041 (43) 国际公布日: 1999年10月14日(14.10.1999)

(21) 国际申请号:

PCT/CN99/00046

(22) 国际申谐日;

1999年4月6日(06.04.1999)

(30) 优先权:

98101156.X 98125015.7 1998年4月6日(06.04.1998) 1998年11月20日(20.11.1998)

CN CN

- (71)(72) 发明人/申请人: 刘莎(LIU, Sha) [CN/CN]; 中国北京市海淀区牡丹园东里10号楼1307室,邮政编码:100083, Beijing (CN)。
- (74) 代理人: 永新专利商标代理有限公司北京办事处(NTD PATENT & TRADEMARK AGENCY LTD., BEIJING OFFICE); 中国北京市金融大街27号投资广场A座10层, 邮政编码:100032, Beijing (CN)。

(81) 指定国:

AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO专利(GH, CM, KE, LS, MW, SD, SL, SZ, UG, ZW), 欧亚专利(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲专利(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI专利(BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG)

本国际公布:

包括国际检索报告。

- (54) Title: OPENING AND HOLOGRAPHIC TEMPLATE TYPE OF LANGUAGE TRANSLATION METHOD HAVING MAN-MACHINE DIALOGUE FUNCTION AND HOLOGRAPHIC SEMANTEME MARKING SYSTEM
- (54) 发明名称: 开放式全息模板式人机对话语言翻译方法及全息语义标注系统

#### (57) Abstract

An opening and holographic template type of language translation method having man-machine dialogue function, includes: creating a natural language restricted dialogue template, wherein it contains all necessary semantic information elements in all of the natural languages; determining vocabulary information items and syntax information items which are actually included in symbols of the natural language through check all type man-machine dialogue on the template; implementing original text solution; generating translation based on the solution; and converting the solution to translation symbols for query of translation syntax. The method makes syntax analysis without depending on the language environment of context and fully utilizes complementary man-machine advantages; this method can be used for eliminating the drawback of syntax information transferring in global network communication.

Sequence	_!_	_2_	3	4	. 5	6	7	8	.7	10
Symbol						١				
			JNF	ORMAT	108	TENS				
Definition of Concepts										
Semantic Semantic									_	
Tense									-	
Vnice	_								1	<u> </u>
Grammar apribuic										
l'an of specch				<b></b> .						
izsciupper position)										LJ

#### (57) 摘要

本发明涉及一种开放式全息模板式人机对话语言翻译方法,包括: 建立一个以句子为对象的自然语言受限对话模板,模板中包含各种自然 语言的所有必要语义信息要素,在模板上通过全选式人机对话确定自然 语言符号实际携带的词汇信息项和句法信息项,完成原文信息求解,根 据求解结果生成译文,并将求解结果转换成译文符号,供译文语义查 询。该方法不依赖上下文语境作语义分析并充分利用人机优势互补的作 用,用于在全球化网络通信中扫除语义信息传递障碍。

以下内容仅供参考 在按照PCT所公布的国际申请小册子首页上所采用的PCT成页国国家代码如下:

					· · · · · · · · · · · · · · · · · · ·		
ΑE	阿拉伯联合哲长四	DK	<b>升</b> 表	КP	胡郎贝主主义人民共和国	PT	葡萄牙
AL	阿尔巴尼亚	DM	<b>步米尼加</b>	KR	56日	RO	PBEW.
AM	亚灵尼亚	EE	爱妙尼亚	ΚZ	哈萨克斯坦	RU	使罗斯联邦
ΤA	<b>奥地</b> 羽	ES	西班牙	LC	<b>至户西亚</b>	SD	<b>が</b> 丹
ΛU	表大利亚	F1	<b>穿</b> 兰	LI	列支敦士亞	SE	<b>始</b> 典
ΑZ	阿世界區	FR	<b>连</b> 园	LK	新知堂卡	SG	新加坡
DA	这所尼亚-孤黎母维那	GA	加蓬	LR	利比里亚	Sī	斯洛文尼亚
BB	巴巴多斯	GB	<b>交</b> 图	LS	<b>莱</b> 安托	SK	斯洛伐克
BE	比利时	CD	格拉纳达	LT	立构建	SL	老拉里品
BF	<b>石灰纳生</b> 素	GE	你会幸业	LU	卢麻堡	SN	窓内加尔
BÇ	<b>保加利亚</b>	CH	to th	LV.	拉托维亚	SZ	斯政士芸
BJ	贝宁	GM	网比亚	MA	摩洛亚	TD	乍视
BR	巴西	GN	几内亚	MC	<b>摩拾</b>	TG	北哥
BY	白保罗斯	GR	<b>安斯</b>	MD	摩尔多瓦共和国	TJ	塔吉克斯坦
CA	加拿大	CW	几内亚比语	MC	乌达加斯加	TM	上床受斯坦
CF	中非共和国	HR	<b>克罗地亚</b>	MK	前叫斯拉夫马其银灰和国	TR	土井瓦
CG	果根	HU	勾牙利	ML	7. T	TT	特立尼达和多巴哥
CH	a±	1D	印度尼西亚	MN	겠삼	17.	坦素尼亚
Cl	科符迪比	Œ	<b>党</b> が立	MR	毛型塔尼亚	UA	QXE
CV)	<b>昭</b> 李陵	IL	以任利	MW	马拉路	UG	马子达
CN	中风	IN	<b>华</b> 皮	MX	<b>歪亚母</b>	US	<b>₩</b>
CR	<b>哥斯达黎加</b>	15	冰岛	NE	足日尔	UZ.	乌兹别克斯却
CU	<b>百</b> 也	IT	龙大科	NL	有三	VN	AS TH
C.Y.	塞加格斯	JP	日本	NO	群城	VU	可斯拉夫
CZ	<b>建克共和国</b>	KE	#EV.	NZ	斯西兰	ZĀ	机非
DΕ	德国	KG	古尔省斯斯坦	PL	改三	ZW	体巴布书
							* * *



# 开放式全息模板式人机对话语言翻译方法及全息语义标注系统

### 技术领域

本发明涉及一种计算机翻译方法,更确切地说是涉及一种在计算机 5 网络中适于各网络终端以不同自然语言进行信息传递交流的机器翻译 方法。

### <u>背景技术</u>

15

20

计算机网络技术以其四通八达、无处不到的优势而迅速开创出一个 全球化的网络信息时代。但由于不同自然语言之间语义信息的传递交流 障碍,己明显制约了网络及网络信息的使用效率,如何通过机器翻译处 理使各网络终端用户仅使用自己的自然语言在网络上进行语义信息传 递,对于节省网络空间、提高网络信息的传递效率和实现网络信息资源 的大众化国际共享,都无疑具有重要的现实意义和很高的商业价值。

目前在机器翻译领域,一方面由人工智能教科书上系统介绍的机器翻译方法在实际的产品开发中很少被使用,另一方面,在己开发出的机器翻译系统中所应用的机器翻译方法又不能达到预期的目标,上述现象说明:基础理论研究严重滞后;所采用的机器翻译技术方法具有普遍共性的缺陷;预期目标本身不具有现实性。进入 90 年代以来,出现了大致两类新兴的机器翻译方法并逐渐成为自然语言信息处理的技术主流。一种是以对大规模真实文本的统计分析为基本手段建设语料库,另一种是人机对话及自然语言受限的机器翻译方法。

大规模真实文本的统计分析是通过对大规模真实文本进行符号、句型、词性、语义等多角度的信息取样分析,从而为任何一种自然语言中的符号与提供多种匹配模式,因而是一种基于经验的语言信息处理方

ı



法。从方法论上说,用这种自然语言信息处理方法可以将源语的多种匹配分析结果作叠加处理,并通过与译出目标语的多种匹配分析结果建立匹配关系,而直接完成自然语言的自动翻译,但现实状况是,自然语言系统具有随机开放特性,任何统计方法都只能提供概率性知识,不可能对自然语言词汇及其概念定义进行准入限制,不可能确定各种省略表达部分的确切内容,也不可能解决生成目标语后的新增歧义。因此,大规模真实文本的统计分析虽然对于利用计算机进行各种自然语言信息处理来说确是有意义的基础工作,但对于机器翻译来说,这种技术手段还需要组合在一种全面有效的对象处理系统方法中才能充分实现其应用价值。

人机对话及自然语言受限的机器翻译方法是由用户在输入端调整 机器翻译词典和调整源语言表达方式,同时调整译文结果,该方法虽可获得较好的机器翻译质量,但要求用户熟练掌握机器翻译的源语和目标语并需付出相当高的人机对话学习成本及操作成本,与人工翻译不相上下。

# 本发明目的

10

15

本发明的目的是设计一种开放式全息模板式人机对话机器翻译方法,以全面解决计算机网络多语种信息传递交流障碍问题,试图取得机20 器翻译技术的实质性突破。这种突破必须满足以下要求:

- 1. 对自然语言普通词汇及其概念定义进行有效的准入限制:
- 2. 不依赖上下文语境进行语义分析;
- 3. 通过直译手段实现语义信息准确传递;
- 4. 找到生成目标语后的新增歧义解决办法;
- 25 5. 用户只需熟练掌握母语:



- 6. 利用大规模真实文本统计分析的手段与成果,充分实现人机优势互补:
  - 7. 满足向多种目标语言转换的需要。

本发明的另一个目的是提出一种全息语义标注系统,利用其可对一 5 文本进行全息语义标注,并将标注信息与文本一起存储。当需要时可将 标注信息与文本一起调出。

### 本发明概述

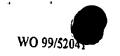
根据本发明的一个方面提出一种开放式全息模板式人机对话语言 10 翻译方法,包括下述步骤:

- a.对各种自然语言进行通约限制;
- b.建立一个以句子为对象的包括各种自然语言的必要语义信息要素的人机对话模板;
- c.由人机对话模板提供与原文语言符号对应的经通约限制的所有备 15 选语义信息项及供用户扩展的空白信息项:
  - d.先由翻译系统的计算机对经通约限制的所有备选语义信息顶进行自动优选,再由原文用户在人机对话模板上对优选结果进行人工调整和确认;
- e.由翻译系统根据人机互补确定的语义信息项生成译文,并将所述 20 人机互补确定的语义信息项转换成译文符号,连同译文提供给译文用户 查询。

根据本发明的另一个方面提出一种全息语义标注系统,包括:

必要语义信息库,其内存有基本词汇及其概念定义以及句法信息 项:

25 文本输入装置,用于输入待对其进行语义标注的文本:



子:

文本存储装置,用于存储通过文本输入装置输入的文本; 文本显示装置,用于显示存储在文本存储装置中的某一文本; 句子选择装置,用于选择文本显示装置所显示的文本中的某一句

5 句子结构自动分析装置,用于根据统计经验自动分析所选句子的结构;

语义标注模板显示装置,用于显示一语义标注模板,该语义标注模板在通过句子选择装置选择了一个句子时对应于所选择的句子被显示,其内包括有对应于该句子中各词汇的词汇信息要素项和句法信息要素项,词汇信息要素项中显示有对应词汇的在必要语义信息库中所包含的概念定义和所有同义词,而各句法信息要素项则根据所述句子结构自动分析装置分析的结果,显示对应词汇的所有可能的句法信息项,该各句法信息项存储于所述必要语义信息库中:

语义标注装置,用于供人对语义标注模板中的各词汇信息要素项中 15 的概念定义及同义词和各句法信息要素项中的句法信息项进行选择;

标注文本存储装置,用于存储带有标注信息的文本;

标注指令装置,用于指令对文本显示装置显示的文本中的某一句子 显示其标注;

标注显示装置,用于以所述标注模板的形式显示存储在标注文本存 20 储装置中的对应于被指令句子的标注信息。

# 工业实用性



域限制;本方法通过建立统一受限标准和全息全选全程的人机对话,提供了一揽子解决机器翻译基本技术障碍的系统方案,为根本改善机器翻译质量提供了全方位的技术保证:本方法可充分利用大规模语料库建设的成果,对自然语言的处理方法简洁实用,具有良好的可实施性;虽然在源语信息求解阶段,用户看不懂的语言不可能进行人机对话,但可在保证翻译质量的前提下实现一种语言输入得到多语种译出结果。

本发明的开放式全息模板式人机对话语言翻译方法在网络信息交流领域具有普遍应用的价值,在打开网络在线机器翻译服务方面有广阔的国际市场。

10 本发明的全息语义标注系统可以将一文本的词汇释义及语法结构信息与文本同时存储,并在需要时显示这些标注信息。该系统可广泛地用于法律文件的释义及语言教学等方面。

# 附图的简要说明

- 15 图 1 是以句子为对象的自然语言全息对话模板结构示意图;
  - 图 2 示出了以一个英文句子为对象的全息对话模板的内容;
  - 图 3 是不同自然语言之间词汇信息通约限制结构示意图;
  - 图 4a、图 4b 是人机对话过程中的两种对话信息显示方法示意图;
  - 图 5 是句法成分信息的空间定位结构示意图;
- 20 图 6 是根据本发明的方法对一个英文句子进行翻译时的人机交互信息处理过程。
  - 图 7 是译文用户查询自然语言符号 "with a telescope" 实际携带的句法信息项的示意图;

# 25 本发明的最佳实施方式



下面结合一个将英文句子翻译成中文的实例说明本发明的开放式全息模板式人机对话语音翻译方法的原理和实施过程。所采用的例句为"I saw a boy with a telescope near the bank."("我在银行附近看见一个带望远镜的男孩。")

5 该例句中包含多个语言符号,这里所说的语言符号既可以是单词,也可以是短语。每个语言符号均携带了一定的语义信息,其中包括语言符号的概念定义、时态、语态以及该语言符号在句子中的成分等多种类别。例如,单词"saw"的概念定义是"看见",时态为过去时,语态为主动语态,在句子中的成分是谓语。但是由于自然语言的复杂多样性,语言符号可能携带不止一个同类语义信息,例如上述单词"saw"的概念除了可以定义为"看见",还可以定义为"理解与认识",又如短语"with a telescope"的句法成分既可以是谓词修饰语,也可以是宾词修饰语。

本发明人认为自然语言翻译的根本任务就是将原文语言符号所携 带的实际语义信息准确地传递给不同语言的用户。为此,本发明采取的 手段是在原文用户方以人机交互的方式对原文的所有语义信息项进行 求解,根据求解的结果生成译文,并将该求解结果转换为译文符号,随 译文提供给译文用户以便查询,从而实现原文用户和译文用户共同参与 的全程翻译,提高语义信息的传递质量。

20 为了对原文的语义信息进行求解,本发明建立了一个如图1所示的 以句子为对象的自然语言全息对话模板,所谓"全息"是指在这个模板 中包括各种自然语言文字符号系统所有必要的语义信息要素,包括属于 词汇信息要素的概念定义项、时态信息项及语态信息项,和属于句法信 息要素的句法成分项。对话模板用来向原文用户提供与原文各语言符号 相应的备选语义信息项以供人机交互选择。这些对话信息项的内容,如

15

20

后面将要说明的,必须受到系统的限制。该对话模板中还包括一些非用户必选信息项,如语义属性、语法属性、上位语义(格)等,这些信息项可以不由用户进行选择,只由计算机进行概率性自动求解,以便为自动转换生成译文提供相关信息。

为了在不同语言之间准确传递语义信息,最好采用直译手段,这是因为机器翻译系统不可能随机调整目标语句子的词汇和句型。但由于各种自然语言的概念体系和句法体系之间存在差异,要想保证直译的译文质量,必须保证词汇信息项和句法信息项能在源语与目标语间作等价交换。因此本发明对不同自然语言间的差异通过建立系统的通约限制原则进行统一整合处理。这种通约限制原则包括句法信息通约和词汇信息通约。

本发明设计的句法信息通约原则包括:统一合并功能同一、对象不同一的句法信息;尽量删除在语义聚合关系分析中并非不可缺少的句法概念,如英语语法中的直接宾语与间接宾语。本发明在对话模板上只提供经简化通约后的句法信息概念,作为不同自然语言的标准句法信息项供用户选择。

本发明设计的词汇信息通约原则如图 3 中所示,是通过对大语种词汇使用频率的统计分析和同义归并而确定一个基础概念集。但实际操作时,不是每一种自然语言的基础概念都是完整对应的,当出现对应空缺时,则要采用该语言的其他常用词汇对这一概念进行解释性描述,使各种语言的基础概念强制性对齐。如英文词汇 orphan 的动词义项被定为基础概念,而中文中没有对应词,则用"使成为孤儿"进行解释性描述。另外,以各种自然语言的基础概念的近义词作为近义附码。由于一种自然语言中某个词汇的全部近义概念也不可能在其它自然语言中全部找到对应概念,因此在当某种自然语言的近义概念出现对应空缺时则由基



15

20

础概念词进行近义替换(人工翻译中近义替换也是不可避免的)。经过上述两项通约处理后仍不能处理的则作为冗余信息在全息对话模板中提供空白信息项。本发明在确定不同自然语言词汇的概念定义时,采用以内涵为中心的模糊通约(如中文的"学校"与英文的"school");不考虑词性差异的概念统一通约(如不考虑英文词汇become 的所有时态变形)和对多种语言中都使用的概念作优先考虑的概率通约处理。为了丰富语言的表达力,任何语言都需要有同一概念的近义词,因此以词汇的使用概率作为词汇概念冗余标准,优先选择多种语言中都使用的词汇,其次是在一种自然语言使用概率高的词汇。对于不满足上述两种情况的词汇则作为冗余概念处理,相应地在全息对话模板中提供空白信息项。经过通约限制处理后的概念定义才作为全息模板中的词汇备选项提供给不同自然语言用户进行选择,以保证不同自然语言词汇概念信息间能够等价互换。本发明还为不同自然语言中相对应的词汇概念设定了统一的编码,以便于在网络上进行信息传递。

另一方面,为了能对系统未收入的自然语言符号进行处理,使得人机交互的方式更加灵活,在通约限制的基本原则下将本发明的对话模板设计成开放式的,即当某个原文自然语言符号未被收入在机器翻译系统中时,原文用户可以调用系统已收入的、己限定信息项的自然语言符号对其进行语义描述。

本发明的对多种自然语言概念系统进行强制性通约受限的方法,与 传统的中间语言方法间有着本质区别:传统的中间语言技术面对的是完 全不受限的自然语言系统,通过建立多种自然语言间的中间概念体系来 实现多语互译,但各种自然语言概念体系的开放性使中间语言体系不可 能具有周延性;强制性的通约受限方法是通过人机对话方式对词汇及义 项作必要的限制和通约,对各种自然语言概念体系之间的差异和开放性



20

进行合理限制,以保证多种自然语言的词汇概念及句法概念能成功地进行等价互换。

PCT/CN99/00046

现在再参见图 2,继续说明原文用户方对原文的语义信息进行求解的方法。该图中示出了人机对话模板向原文用户提供的与原文各语言符号相应的经通约限制的备选语义信息项。对原文的语义信息进行求解的过程也就是对人机对话模板中的这些备选信息项的选择、确认和补充的过程。

在对词汇信息项的选择中要充分利用人机优势互补,计算机自动优选所遵循的基本原则是:通过大规模的对真实文本的统计分析,排列出多义词的词汇信息项使用频率顺序,以缩小用户选项的搜寻范围;通过大规模的对真实文本的统计分析,根据句法信息项与词汇信息项间的相关性特性来优选词汇信息项,以进一步缩小信息项选择范围,例如凡可做主语的词汇都优选其名词义项,象图 2 中的"I"和"telescope"等;通过大规模的对真实文本的统计分析,获得词汇搭配的概率信息,进一步优选词汇信息项,如汉语"好漂亮的一朵花",其中的"好"是多义词,而在形容词"漂亮"前的"好"字的最可能的义项解是程度副词"非常";对于显性表达词性信息的文字符号,通过词性即可推导出所选词汇信息项来缩小信息项选择范围,如英语中"spring"的词根虽然是多义的,但其动词的过去式"sprang"则已明确限制了义项选择范围。

通过以上技术手段的自动选项处理,己能够将用户实际所需的大多数词汇信息项排在首位,由于表达语义所需要的词汇信息项已存在于用户心中,因此对用户而言,大多数的词汇信息项选择只是一个对模板中各首选信息项的确认过程。

各种自然语言中,无论是隐性表达还是显性表达的句法信息,大体 25 上包括词性信息、句法成分信息和上位语义(格)信息,其中句法成分



20

25

信息是唯一具有完整组织能力的,并具有普遍共性的句法组织系统,因 此,只要确定句法成分信息项,实际上已经确定了一个自然语言符号串 的语义聚合关系。在句法信息项的选择中也要充分利用人机优势互补, 其所遵循的基本原则是: 通过大规模的对真实文本的统计分析获得词 序、词性、上位语义(格)信息与句法信息之间的匹配关系,以自动优 选句法信息项。如一个词汇的词序为1,词性为名词,上位语义为行为 主体,则可判定为主语:用户通过选项操作最终确定句法成分信息项。

通过以人机对话方式在模板上选定词汇信息项和句法信息项, 求解 原文的实际语义信息。由用户直接在全息对话模板上选择各自然语言符 号串实际携带的词汇信息项和句法信息项,是最简单的人机对话方式, 其具体方法可以是对所确定的项进行黑体标注处理,如图〔中所示。

通过在全息对话模板中对句子中词汇信息项和句法信息项的人机 互补选择、确认,已能够完成自然语言的信息求解任务,因此不再需要 依赖上下文语境对句子进行语义分析,

对于用户来说,分析和确定抽象的句法关系远比判断多义词信息项 15′ ` 困难,因此,为了降低句法成分信息项的选择难度,实际操作时可象图 5 中所示的那样将呈线性排列的句法成分信息项转换成空间定位表达方 式, 协助进行句法成分信息项人机对话的选择。以句法信息的修饰区、 核心区及补充区为横座标,以句法信息的主语区、谓语区及宾语区为纵 座标,作出句法信息对话框架,由用户在框架中对"with a telescope" 的修饰对象进行选择。

在实际的人机对话过程中也可以采用模板部分显示方法和模板虚 拟方法,如图 4a 所示的句法信息全显(图中?号表示由用户再选择)和图 4b 所示的 "I see a boy with a telescope near the bank" 的虚拟对话模板后 的对话显示方法。本领域的技术人员应该理解,人机对话过程中的对话



10

15

20

信息显示方法可以有很多种, 而不限于本说明书中的示例。

本发明的方法通过对语法概念和普通概念的系统通约受限,以及在 受限信息项范围内进行人机互补信息全选,已经具有了向多种自然语言 表达形式作自动转换的必要信息,但总有被用户省略的句法成分,从逻 辑上说只要确定了已有文字符号的所有信息项,大多数省略部分可由用 户在阅读信息时根据上下文语境自动添加(如主词、宾词省略),但为了 准确传递语义,对不可省略的句子成分还要通过全息对话模板进行添 加,以保证机器翻译质量(如在一个句子的备选信息项中已经选了主词 和宾词,则不可省略相关动词)。

为了解决生成目标语译文后发现新增歧义的问题,将经过全息对话的中间翻译结果随译文提供给目标语用户作直接查询,可实现目标语新增歧义的全面消解。如果用户有意保留语言表达的模糊性或双关性,则可在选择信息项时作多项同时选择。

参见图 6,图中流程说明了本发明的开放式全息模板式人机对话语言翻译方法中的人机交互信息处理基本过程,其中中间列框 11 至 17 是翻译系统计算机的主流程,左边列框 21 至 26 示出用户的参与过程,右边列框 31 至 35 示出人机交互过程中与内部数据库、规则库间的关系,单向箭头表示人机交互流向,双向箭头表示在语言翻译过程中对数据及规则的调用过程,所标的 N 表示系统信息处理需要人机交互,所标的 Y 表示自动进入系统流程的下一个操作步骤, # # - - - - # # 表示此翻译系统与因特网系统的信息处理接口。其上方表示原文用户端,其下方表示译文用户端。

处理过程开始, 执行步骤 11. 由原文用户顺序输入待翻译的自然语言符号。

25 结合参见图 2, 在模板的序位 1 至 10 中依次填入本例中的十个自然

WO 99/52041

the bank"; a boy with a telescope near 语言符号"I saw 系统主程序的步骤 12 在可扩展的多语对应的词汇信息项符号库 31 中, 对各自然语言符号进行词汇备用信息项搜索, 当搜索不到时可通过步骤 21 由原文用户在模板上用系统已收入的语义符号对自然语言符号的语 义进行描述,上述过程最终生成模板中的由概念定义项、语义属性项、 5 时态项、语态项等构成的词汇备用信息项,如果在某自然语言符号下出 现概念定义信息项空白,如在符号"bank"处出现"?",则原文用户 可采用系统中已提供有信息项的词汇符号对其进行语义描述,即模板中 概念定义项 "institution for keeping or lending money"; 系统主程序的 步骤 13, 根据词汇信息项概率性优选规则库 32 中的规则, 由计算机对 10 列入模板中的各自然语言符号的多个词汇备用信息项进行自动优选,如 模板中用黑体字指定的信息项,并可通过步骤 22 由原文用户对未获得 确定优选的语义信息项进行选择确认;系统主程序的步骤 14,通过调 用句法成分信息项自动标注规则库33,对列入模板中的各自然语言符 号的句法信息项进行自动标注,上述过程最终生成模板中的句法成分 15 项、词性项、上位"格"项;系统主程序的步骤15,调用句法成分信息 项自动优选规则库 34, 对各自然语言符号的句法成分信息项进行自动 优选, 其间可通过步骤 24 调用句法信息项三维结构模型库 23, 由原文 用户在模板上对未获得唯一优选结果的句法信息项进行选择确认,如模 板中用黑体字指定的信息项; 系统主程序至此就可以以自定的编码形式 20 在网络上传递所确定的上述信息项。

对话模板中包括了自然语言符号所能携带的所有信息项,其全部备用信息项不仅包括自然语言符号的概念定义、时态信息、语态信息、句法信息、上位"格"信息、词性信息、单复数信息、阴阳性信息、而且可在开放式的模板下部扩展人工设计和标注的其它信息。



15

20

当原文用户在图 6 的步骤 21 中利用语义描述方法求解原文符号时,系统程序还要自动进行其使用频率的统计,在使用频率达到一定水平时,即在翻译系统收入的所有语种的自然语言符号库中同步增添新增的自然语言符号或新增信息项。如人工描述求解 bank 的使用频率达到一定水平时,系统即在法语的自然语言符号库中添加新增的符号"banque"和利用系统已收入的相应法语符号进行语义进步,并依如此

"banque"和利用系统已收入的相应法语符号进行语义描述,并给出其它相关备选信息项。其它语言的扩展方法与其相同。

结合参见图 7,译文用户端系统主程序的步骤 16,调用译文自动转换生成规则库 35,根据多语符号与序位转换规则,将经原文用户确认的信息项求解结果自动转换成译文用户要求的自然语言译文,如图 7中所示的中文转换生成结果"在银行附近我看见一男孩带望远镜";系统主程序在步骤 17 将会询问用户译文是否已经无歧义,如果有歧义,译文用户可在步骤 26 通过人机交互过程确定相关信息项的查询范围,其间可调用多语对应信息项符号库 25,如译文用户为了求解"带望远镜"到底是修饰主语还是修饰宾语,如图 7中?所示,就可直接查询该符号实际携带的句法信息项,从而确定是修饰宾语。至此翻译过程结束。

语义信息传递质量是全球化网络信息时代机器翻译技术赢得巨大国际市场的根本障碍,要想取得实质性突破,人机对话是不可避免的,本发明人机对话优势互补的翻译方案可切实提高翻译质量,具有实用价值。由于本方法具有语义信息传递准确、不受语言环境限制、用户操作使用方便、可同步转换生成多种目标语、对话方案多语通用及技术手段简单可靠等优点,因而在网络信息交流领域将会具有普遍应用价值,在网络的在线机译服务方面也会有广阔的市场。

根据上述方法的构思本发明还提供了一种全息语义标注系统,该系 25 统包括:



必要语义信息库,其内存有基本词汇及其概念定义以及句法信息 项;

文本输入装置,用于输入待对其进行语义标注的文本:

文本存储装置,用于存储通过文本输入装置输入的文本;

5 文本显示装置,用于显示存储在文本存储装置中的某一文本:

句子选择装置,用于选择文本显示装置所显示的文本中的某一句子:

句子结构自动分析装置,用于根据统计经验自动分析所选句子的结构;

10 语义标注模板显示装置,用于显示一语义标注模板,该语义标注模板在通过句子选择装置选择了一个句子时对应于所选择的句子被显示,其内包括有对应于该句子中各词汇的词汇信息要素项和句法信息要素项,词汇信息要素项中显示有对应词汇的在必要语义信息库中所包含的概念定义和所有同义词,而各句法信息要素项则根据所述句子结构自动分析装置分析的结果,显示对应词汇的所有可能的句法信息项,该各句法信息项存储于所述必要语义信息库中。

语义标注装置,用于供人对语义标注模板中的各词汇信息要素项中的概念定义及同义词和各句法信息要素项中的句法信息项进行选择;

标注文本存储装置,用于存储带有标注信息的文本;

20 标注指令装置,用于指令对文本显示装置显示的文本中的某一句子 显示其标注;

标注显示装置,用于以所述标注模板的形式显示存储在标注文本存储装置中的对应于被指令句子的标注信息。

25 本发明的该全息语义标注系统的一种应用为同文种全息语义标注系统,以法律业为例: 法律分门别类很多,需要建立相应的知识库。开发专家系统具有广泛的应用价值。其中一个带普遍性的应用需求是普通用户对



法律条文的语义理解和识别。国内外已有的各种专家系统,都是"问答式"人-机介面:系统依次提出许许多多问题,用户逐一作出"Yes"或"No"的选择,或者输入简单的数据,然后由系统搜索知识库,根据问题与知识的匹配情况推断出某个结论,然后告诉用户。

5 这种"问答式"人-机介面呆板、繁琐,并且系统所提问题事先设定, 不灵活。这样的系统显得智商太低。

如果在输入法律解释条文、合同、协议、诉状时,采用同文种语义标注技术,一次输入所使用语言符号的全息数据,将大大方便用户查询及分类整理。

10 同文种语义标注技术不仅适用于建立各类专家级知识系统开发,而且 对于提高法律解释、合同内容、技术说明文件的语义表述精确性均具有普 遍实用价值。

同文种语义标注技术的实现方法:

仅应用全息翻译模板的原文加工技术并提供专业词库,即可实现同文 15 种语义标注。

本发明的该全息语义标注系统的一种应用为外语全息语言教学系统。

计算机辅助教学目前已经应用的十分广泛。在外语教学领域领域的应用主要采用的是多媒体教学法(听、说、读、写并行)和应试题库教学。而语言全息模板为外语教学提供了一种系统反映不同语言概念共性和符号化个性的计算机辅助教学手段。

当用户输入母语句子时:

20

如果用户选定母语词汇的概念定义,通过系统提供的多语统一编码,全息模板即可调出多种语言的所有对应词汇。

25 如果用户选定母语句子的时态、语态、句法成分信息项,全息教学





系统则可利用全息翻译系统的界面技术及内部转换规则,分步提供任意语种的符号变形和排序变换过程。

如果用户直接输入外语句子,通过系统提供的多语统一编码,全息 模板则既可提供外语的全息语义标注,也可将全息语义标注直接转换为 母语。



## 权利要求书

- 1.一种开放式全息模板式人机对话语言翻译方法, 其特征在于包括下 述步骤:
- a.对各种自然语言进行通约限制; 5
  - b.建立一个以句子为对象的包括各种自然语言的必要语义信息要素 的人机对话模板;
  - c.由人机对话模板提供与原文语言符号对应的经通约限制的所有备 选语义信息项及供用户扩展的空白信息项;
- d.先由翻译系统的计算机对经通约限制的所有备选语义信息顶进行 10 自动优选,再由原文用户在人机对话模板上对优选结果进行人工调整和 确认:
- e.由翻译系统根据人机互补确定的语义信息项生成译文,并将所述人 机互补确定的语义信息项转换成译文符号,连同译文提供给译文用户查 15 询。
  - 2.根据权利要求1所述的开放式全息模板式人机对话语言翻译方 法, 其特征在于: 所述步骤 b 中的必要语义信息要素包括概念定义、时 态信息、语态信息和句法成分信息项。
- 3.根据权利要求1或2所述的开放式全息模板式人机对话语言翻译方 法, 其特征在于所述步骤 a 中的对各种自然语言的通约限制包括: al. 20 统一合并功能同一、对象不同一的句法概念; a2.尽量删除可缺少的句法 概念: a3.通过对主要语种词汇使用频率的统计分析和同义归并, 建立多 语通用基础概念集; a4 以各种自然语言的基础概念的近义词作为近义附 码,当不同自然语言出现近义词对应空缺时,由基础概念词进行近义替 换: a5.对于不能用基础概念进行统一表达的自然语言词汇或概念,由对 25



话模板提供空白信息项。

4.根据权利要求 1 所述的开放式全息模板式人机对话语言翻译方法, 其特征在于: 所述的步骤 c 中, 当与原文语言符号项对应的同语种某备选信息项中出现有空白时, 用户可调用己被系统收入的自然语言符号对其进行描述。

5.根据权利要求 4 所述的开放式全息模板式人机对话语言翻译方法, 其特征在于该方法还包括: 对由用户扩展的信息项进行使用频率统计, 并根据使用频率统计结果确定新增通用基础概念, 在翻译系统的所有语种的人机对话模板中同步增添自然语言符号项及对应的信息项。

10 6. 根据权利要求 1 所述的开放式全息模板式人机对话语言翻译方法, 其特征在于: 所述步骤 d 的对自动优选结果进行人工调整和确认的方法是由用户在全息对话模板上对不确定信息项进行人工选择。

7.根据权利要求1所述的开放式全息模板式人机对话语言翻译方法,其特征在于:所述步骤b的以句子为对象的人机对话模板是包括有三维空间定位句法的对话框架。

8.根据权利要求 1 所述的开放式全息模板式人机对话语言翻译方法, 其特征在于: 所述步骤 b 的以句子为对象的人机对话模板是虚拟的。

9.根据权利要求 3 所述的开放式全息模板式人机对话语言翻译方 20 法,其特征在于:对各种自然语言的通约限制方法还包括有 a6.以内涵 为中心的模糊通约和 a7.不考虑词性差异的概念统一通约。

10.根据权利要求1所述的开放式全息模板式人机对话语言翻译方法,其特征在于:所述步骤d中,用户可单项或多项地在全息对话模板上对优选结果进行人工调整和确认选择。



11. 一种全息语义标注系统,包括:

必要语义信息库,其内存有基本词汇及其概念定义以及句法信息 项;

文本输入装置,用于输入待对其进行语义标注的文本;

5 文本存储装置,用于存储通过文本输入装置输入的文本;

文本显示装置,用于显示存储在文本存储装置中的某一文本;

句子选择装置,用于选择文本显示装置所显示的文本中的某一句子:

句子结构自动分析装置,用于根据统计经验自动分析所选句子的结 10 构;

语义标注模板显示装置,用于显示一语义标注模板,该语义标注模板在通过句子选择装置选择了一个句子时对应于所选择的句子被显示,其内包括有对应于该句子中各词汇的词汇信息要素项和句法信息要素项,词汇信息要素项中显示有对应词汇的在必要语义信息库中所包含的概念定义和所有同义词,而各句法信息要素项则根据所述句子结构自动分析装置分析的结果,显示对应词汇的所有可能的句法信息项,该各句法信息项存储于所述必要语义信息库中:

语义标注装置,用于供人对语义标注模板中的各词汇信息要素项中的概念定义及同义词和各句法信息要素项中的句法信息项进行选择:

20 标注文本存储装置,用于存储带有标注信息的文本:

标注指令装置,用于指令对文本显示装置显示的文本中的某一句子显示其标注;

标注显示装置,用于以所述标注模板的形式显示存储在标注文本存储装置中的对应于被指令句子的标注信息。

25 12. 根据权利要求 11 所述的全息语义标注系统, 其特征在于所述必要语义信息库中相对应地存有多种语种的通约受限的词汇及其概念定义并且相对应地存有多种语种的通约受限的句法信息项。



• 5

- 13. 根据权利要求 11 所述的全息语义标注系统,其特征在于在所述的某词汇的词汇信息要素项中还显示必要语义信息库中与该词汇对应存储的指定语种的词汇且该词汇的句法信息要素项中还显示必要语义信息库中与该词汇的句法信息项相对应地存储的该指定语种的句法信息项。
  - 14. 根据权利要求 11 所述的全息语义标注系统, 其特征在于在所述的词汇信息要素项中的内容除所述可供选择的内容外, 还可被改为其他用于说明该词汇意义的信息。



Sequence	1	2	3	4	5	6	7	8	9	10	Τ
Symbol									1	-5	Ħ
	•		IN	ORMAT	ION	ITEMS		·	- <del>L</del>	-4	
Definition of Concepts											
Semantic attributes											<del> -</del>
Tense		·			<b></b>						
Voice	-								ļ <u>.</u>		
Grammar attribute					~						
Part of speech								-			
Case(upper position)											

Sequence	1	2		3	4	5	6	7	8	9	10	
Symbol	I	saw		a	boy	with	a	telescope	near	the	bank	Ţ.
· · · · · · · · · · · · · · · · · · ·			1		ΊN	FORMAT	ION I	TEMS				-
	1.the	to use	1.0	one	A	l. in	l.one	a	1.not far	l.specific	?:	Γ
Definition	person	the	2.1	the	youn	the	2.the	tubelike	from in	person or	instituti	l
of	speaki	eyes	fi:	rst		presen	first	scientific	distance	thing.	on for	
Concepts	ng	to		ter	male	ce of	letter	instrume	, time,	2.а дтоцр	keeping	-
	2. the	look		ıf	pers	2.beside		nt, used	•••••	of	or	
	9th	at;		glish		3.among				3.used for	lending	
	letter of	to	alp	hab		4.inclu	alphab	making		comparas	money	
	the	under	•	et .		ding	et	distant		on	İ	
	English	stand				`		objects				
	alphabe	or						look			ł	
	ί	recog						larger or		j		
		nize	<u> </u>					closer			<u> </u>	L
Semantic	countab		ma	le;	male;		}	countable;			countab	ļ
attributes	le;				count	1		single			le;	
i	single		ble	•	able;		]		1.	1	single	
			sin	gle	single	:						L
Tense	*	past										l
		tense									-	<u> </u>
Voice		active										
		passiv					1					
		е						<u> </u>	<u> </u>			↓_
	Бьоиоп	1 1	ı	ici	noun	prepo			Preposit		noun;	
attribute	n		e,			sition	indefi	ı verb	ion.	adverb	verb	
			ind			ļ	ite		verb.		·····	
			nit			ļ	<u> </u>		adverb.	l	<u> </u>	ļ
Part of	subject	predi						e predicate			the	
speech		cate		of	t	1	odifier	of the	predica	te		
			i	1e		object .	?					
	<u> </u>		ob]	ect		ļ	···	<del></del>				Ļ_
Case(upper	agent	oehavi			bject			tool;		1.	place	
position)		or of	۲	1	f the			behavior			behavior	
	{	act	1	bel	avior	1	1					

		统一编码	英语	汉语	俄语	日语	<u></u>
李	础词	01	Laugh	笑			
	A、程度+	01-1-x	Burst into laughter	大笑			
	B. 程度-	01-2-x	Smile	微笑			
近义附码	C. 成语	01-3-x		笑逐颜开			
	D. 书面语	01-4-x					
	E. 口语	01-5-x		英哈哈			
	F. 俚语	01-6-x					
	G. 俗语	01-7-x					
	H. 褒义	01-8-x					
	I. 贬义	01-9-x	Snicker	傻笑			
	其他近义	01-10-x	Grin	咧嘴笑			

I	saw	a	boy	with	a	telescope	пеат	the	bank .
subject	predicate		Object		pre lific	ation of the dicate ation of the ject	I	oredic	on of the

图 4a

"I see a boy with a telescope [ "modification of the subject" OR

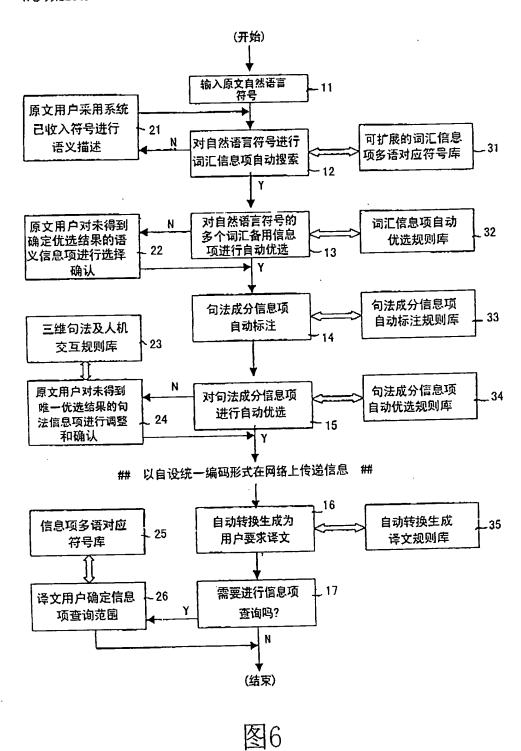
"modification of the predicate" ] near the bank." [ "institution for
keeping or lending money" OR "edge of a river or canal" OR "long
mound of earth/sand/snow" ]

图 4b



	修饰区	核心区	补充区
主语区	.∵·	I	
谓语区	? with a telescope	see	near the bank
宾语区	? with a telescope	boy	
L	整句连词	<u> </u>	标点





6/7

										<del> ,</del>	
符号	在	银行	附近	我	看见		男孩	带	望远镜	的	}
			L							<del>  </del>	
						信息项				<del></del>	$\vdash$
词汇								携带			
信息			1					? 附带	L		L
句法	1	修饰主i	j.	主词	谓词		宾词	修	饰宾词		
رے ۔۔۔ ر					1		İ	ļ			ı

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN 99/00046

		161/6	11 337 1110 10
A. CLASSI	PICATION OF SUBJECT MATTER		
	G06F I	7/28	
According to i	nternational Patent Classification (IPC) or to both national	classification and IPC	
	SEARCHED	(Figure 2)	
Minimum doc	unrentation searched (classification system followed by class		
	G06F 17/28. 17/27. 17/20. 17		
Documentation	searched other than minimum documentation to the extent	that such documents are included in the fi	ieids serrenal
Electronic data	a base consulted during the International search (name of d	ata base and, where practicable, search to	rms used)
Electronic on	a Last Colonises doing the principles		
C. DOCUM	IENTENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where ap	propriate, of the relevant passages	Relevant to claim No.
Category	Charlot of december		-
A	US 5418717A The whole document		1—14
	EP 0622742A1 The whole document		1—14
A	EF 00221 12/11 The whole township		
			1—14
Α	US 5285386A The whole document		,
A	CN 87106964A The whole documnet		1—14
[]] Furthe	r documents are listed in the continuation of Box C		
" A" documer ered to a tional fit documer which is or other er mean "P" documer mean "P" documer documer mean "P" documer mean "P" documer er mean "P" documer er mean "P" documer ered ered "P" documer ered ered ered ered ered ered ered e	nt defining the general state of the art which is not consid- be of particular relevance pplication or patent but published on or after the interna- ling date int which may throw doubts on priority claims (s) or s cited to establish the publication date of another citation or special reason (as specified) intereferring to an oral disclosure, use, exhibition or oth-	"T" later document published after the interty date and not in conflict with the assand the principle or theory underlying."  "X" document of particular relevance: the considered novel or cannot be considered to involve an inventive step interest of involve an inventive step bined with one or more other such decing obvious to a person skilled in the "&" document member of the same patent	application but cited to under- ing the invention cannot be leted to involve an inventive e e chained invention cannot be ep when the document is com- americs, such combination be- art family
Date of the	actual completion of the international search	Date of nucling of the internationa	
	08 June 1999 (08, 06, 99)	0 1 JUL 1999 (0 1. 0	1, 9 9)
Name and	mailing address of the ISA/CN	Authorized officer	
6 Xitucher Beijing Cl	ng Rd. Jimen Bridge, Haidian District, 10008R hina	Guotiong. X	(in
Foreignile !	No. 86 10-62019451	Telephone No. 86-10-62093837	·

# INTERNATIONAL SEARCH REPORT

Information patent family members

International application No-PCT/CN 99/00046

		the state of the s	
Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5418717A	23. 05. 95	CN 1073290A	16. 06. 93
EP 0622742A1	02.11.94	JP 6314294A	08. 11. 94
US 5285386A	08. 02. 91	¹P 3202954A	04. 09 91
		CN 1053308A	24. 07. 91
		CN 1030871B	31. 01. 96
CN 87106964A	01.06.88	None	

## 国际检索报告

国际中语号

PCT/CN 99/00046

. 主題的分			
<b>を照国际者</b>	利分类表 (IPC) 或者同时按照国家分类和 IPC 两种分	类 ————————————————————————————————————	
1. 检索领域	ž.		· .
检索的最份	、限度文献(标明分类体系和分类号) GOGF 17/28, 17/27, 17/20, 17	7/00 . 15/38 <sup>a</sup> · . J5/20 <sup>a</sup> · . 15/00	
包含在检索	\$领域中的除最低限度文献以外的检索文献		
在国际检索	<b>长时在阅的电子数据库(数据库的名称和,如果实际可</b>	行的,使用的检查词)	
c 相关文	件		- Charles
类 型*	引用文件、必要时、包括相关的	2落的说明	相关的权利要求编号
A	US 5418717A 全文		1—14
۸	EP 0622742∧1 全文	·	114
٨	! (IS 528578GA 全文		!—14
٨	(N 8710696-1A 全文		ll4
□ 共介	文件在C栏的线页中列出。	区 见同族专利附件。	
"A" ( "E" ( "L" 7	文件的专用类型: 明确表示了一般现存技术、不认为是特别相关的文件 生国际申请目的当天或之后公布的在先的申请或专利 讨优先权要求可能产生怀疑或者用亲确定另一篇引用 文件的公布日切或其它特殊理由而引用的文件(如详 也说明) 涉及口头公开、使用、展览或其它方式公开的文件 在国际中谓日之前但迟于所要求的优先权日公布的文件	"X"特别相关的文件; 当该文件; 不能认为是新疆的或不能认 "2"特别相关的文件。当该文件。	理解构成发明共2000度比较低度 战单处使用时、要求保护的发明 为具有创造性 与其它一路成多路过考文件结合 技术人员是显而易见的。要求信
国际检·	京实际完成的日期 08. 6月. 1999 (02. 06. 99)		(01.07.99)
}	中国北京市道湾区将门桥西土城路6号	受权官员 夏国主 电话号码: 86-10-62093837	- 12 12 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

国际检索报告 同族专利成员的情报

国际中语号

PCT/CN 99/00046

位素报告中引用的 专利文件	<b>公布日期</b>	同联专利成员	公布日期
US 5418717A	23. 05. 95	CN 1073290A	16. 06. 93
EP 0622742^1	02. 11. 94	JP 6314294A	08. 11. 94
US 5285386A	08. 02. 94	Љ 3202954∧	04. 09. 91
		CN 1053308A	24. 07. 91
		CN 103087 [II	31. 01. 96
CN 87106964A	01. 06. 88	₹.	